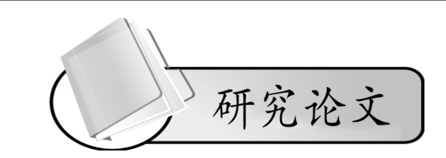


基于相关性的跨模态信息检索研究*

丁　恒[1]　陆　伟[1,2]

[1](武汉大学信息管理学院　武汉　430072)

[2](武汉大学信息资源研究中心　武汉　430072)

摘要:【目的】梳理基于相关性的跨模态信息检索中的基本策略和核心问题，从提升检索效果的角度探讨偏最小二乘法用于特征子空间投影的优劣。**【方法】**在 Wikipedia 跨模态信息检索数据集上，分别采用 LDA 和 BOW 模型作为文本和图像资源的特征表达方式，以余弦距离作为相似度度量方法，利用最小二乘法替代典型相关性分析法学习特征子空间投影函数。**【结果】**从 P@K、MAP 和 NDCG 三个检索评价指标上，对比分析典型相关性分析、偏最小二乘回归、偏最小二乘相关三种特征子空间投影法对跨模态信息检索结果的影响，结果表明偏最小二乘相关法具有最佳效果。**【局限】**偏最小二乘法在处理数据时假设数据之间的关系是线性的，数据基向量之间是正交关系，因而无法解决非线性、非正交问题。**【结论】**使用偏最小二乘相关法学习的特征子空间投影与原始空间信息的一致性更强，跨模态信息检索结果更稳定。

关键词: 跨模态信息检索　偏最小二乘法　子空间投影

分类号: TP393

1　引　言

随着多媒体技术的进步和发展，信息资源多元化程度日益加深，推动了传统信息检索技术的巨大变革，传统的基于文本的信息检索技术逐步向基于内容的多媒体信息检索发展。诸如基于内容的图像检索[1]、基于指纹的音乐检索[2]、基于内容的视频检索[3]等多媒体信息检索研究日益成熟，出现了“以图搜图”、“哼唱检索”等相关商业应用，一定程度上解决了同形态空间内信息资源的检索问题。然而，有时检索系统会面临如下需求：“用户有一张鸟的照片，希望查找到其相关的文字介绍，以及视频和音频片段”，该类检索可以归结为“如何解决不同形态空间(文本、视频、音频、图像等)之间信息资源的相互检索”这一问题。

目前，信息检索系统多利用基于内容的多媒体检索技术，通过查找同形态空间下的相关信息资源，整合这些相关信息资源的目标形态空间信息，最终返回相应结果列表。例如“互联网以图搜文”时，首先通过“以图搜图”查找包含相似图像的网页，然后返回网页中与相似图像相关的文本信息。该方法存在两个重要缺陷：无法检索到不包含图像的网页中的相关文本信息；网页中与相似图像相关的文本信息并不一定与查询图像相关。跨模态信息检索试图直接建立信息资源在不同形态空间内的关联关系，以弥补上述缺陷。

跨模态信息检索(也称跨媒体信息检索)是多媒体信息检索中一个较新的研究领域，涉及到多媒体信息表达、异构特征关联挖掘、子空间投影、语义推理等相关技术，其通过建立信息在多种形态之间的映射，实现信息在不同形态空间中的表达转化，最终支持跨越信息资源形态差异(异构数据类型)的检索。本文通过设计跨模态信息检索实验，以三种常用的信息检索评价指标为基准，探索了不同的多元统计分析方法处

通讯作者：陆伟，ORCID：0000-0002-0929-7416，E-mail：weilu@whu.edu.cn。

*本文系国家自然科学基金面上项目“基于语言模型的通用实体检索建模及框架实现研究”(项目编号:71173164)的研究成果之一。

理异构特征信息，进行特征子空间投影的优点和缺点。本文主要贡献与创新在于，一方面梳理并归纳了基于相关性的跨模态信息研究的核心步骤和策略，另一方面针对子空间投影步骤，提出了以偏最小二乘法挖掘异构特征关联关系的思路，并通过实验结果证实了偏最小二乘法与传统的典型相关性分析法相比，更适用于基于相关性的跨模态信息检索框架。

2 相关研究

2.1 多媒体信息处理

多媒体信息处理技术已经在很多研究领域得到广泛应用，如文献[4]将图像局部不变特征聚类成视觉词汇，并采用空间金字塔模型将图像区域语义信息与“词袋模型”结合起来，实现了对图像场景语义的分析和理解。文献[5]运用隐含狄列克雷分配(Latent Dirichlet Allocation)对短文本进行建模，同时考虑短文本的特征稀疏性和上下文依赖性，从主题层次探讨了短文本的语义理解问题。文献[6]将层次狄列克雷过程(Hierarchical Dirichlet Process)运用在搜索引擎的用户日志分析上，通过对查询词中的动词及与动词具有依存关系的名词进行聚类，进而解决了用户查询意图的语义理解问题。文献[2]在信号频谱分析的基础上，使用快速组合哈希(Fast Combinatorial Hashing)算法对音乐进行信息建模，实现了基于“音乐指纹”的音频信息检索。这些多媒体信息处理技术为同构信息的检索、推荐等应用提供了可能，且一定程度上能够表征信息资源的语义内涵。

2.2 跨媒体语义信息挖掘

然而，多媒体信息处理技术未能在信息资源的异构特征之间架起桥梁，因此一些研究者在此基础上探索性地研究了跨媒体信息之间的内在联系。文献[7]指出同一信息资源在不同形态下的特征之间存在某种潜在联系，并利用典型相关性分析法对这种异构数据(音频与图像)之间的关联关系进行建模，从而将不同形态的信息资源转化到同一子空间中，进而实现了音频与图像之间的跨媒体信息衡量。文献[8]提出运用奇异值分解和隐性语义索引对跨媒体信息进行语义关系建模的思路，并通过跨媒体检索试验对比分析奇异值分解、隐性语义索引和典型相关性分析在异构特征关系挖掘上的优劣。文献[9-10]在跨媒体信息处理过程中引入本体技术，通过基于关系的知识推理和本体学习，构建多媒体信息间的语义关联，从而衡量跨媒体间的信息差异。文献[11]论述了图像和音频内容表达一致性问题，提出一种半监督式的相关性保持映射算法(SSCPM)，用以挖掘图像和音频数据特征之间的潜在共性。文献[12-13]通过检索实验的评价结果，分析多媒体低层特征和高层语义特征在跨模态信息检索任务上的差异，并指出多层次的特征融合更能表征信息在跨媒体数据中的共性。文献[14-15]分别提出时空上下文语义机模型和邻近图模型，从信息资源的跨模态相关性投影(Cross-Modality Correlation Propagation)讨论了跨媒体语义信息的挖掘方法，并探究了文本与图像信息的相互检索问题。这些跨媒体语义信息挖掘研究的主要思路是“构建一个同形语义子空间，对不同维度、不同量纲的特征数据进行空间投影，从而实现对跨模态信息的关系度量，以服务跨模态信息检索研究”。如何学习到一个既保持多模态信息个体差异特性又融合多模态信息共性的子空间是该类研究的核心问题。当前的同形特征子空间构建可归为以下两类方法:

(1) 基于相关性的特征子空间投影：该方法依据最大相关性策略，多采用典型相关性分析法，挖掘不同模态信息低层特征之间的潜在相关关系，学习最优子空间投影矩阵，以实现异构特征空间转换。

(2) 基于高层语义的特征子空间学习：该方法多利用机器学习方法，通过分类算法直接在语义层次上为异构数据构建同形语义特征空间，并基于该空间实现异构数据的相似度度量。

其中第二类方法严重依靠多类分类算法的效果。然而，随着分类类别的增加，多类分类效果往往呈递减趋势，这限制了可构建的语义特征空间的维度，本质上降低了检索对象之间的区分度。其次，该类方法下，若拓展语义特征空间的维度，则需重新学习分类模型、调优参数，是一种参数相关的解决方法，难以适用于实际的检索应用，因此本文仅探讨“基于相关性的特征子空间投影”的跨模态信息检索优化问题。

3 基于相关性的跨模态信息检索

笔者认为基于相关性的跨模态信息检索系统框架主要由多模态信息表达、特征子空间投影和相似度度量排序三个部分组成。

3.1　多模态信息特征表达

多模态信息表达主要研究同形态下信息资源如何编码以便于有效区别类内的个体差异。形式上可直观地认为，多模态信息表达就是利用数学向量从不同角度刻画信息资源本身，其不同角度体现在同一信息资源可使用不同维度、不同数值的向量表示。信息资源在某一特定形态下的特征表达，可用如下形式化定义描述:

对于给定的信息资源集合 $S=\{S_1,S_2,\cdots,S_k\}$，找到一个 m 维的向量空间 L，使得每个信息资源 S_i 在该空间中都可用某一向量 $S_i=\{L_{S_i}^1,L_{S_i}^2,\cdots,L_{S_i}^m\}$ 表示。本文使用 LDA 主题空间和 BOW 视觉词袋空间分别作为信息资源的文本特征表达和视觉特征表达。

3.2　基于相关性的特征子空间投影

特征子空间投影是指在不同形态特征空间下，分析信息资源异构特征之间的潜在联系，从而将异构数据投影到同一特征子空间内，以解决特征异构的问题。基于相关性的特征子空间投影是挖掘不同模态信息低层特征之间的潜在相关关系，学习最优子空间投影矩阵，以实现异构特征空间转换，其核心在于将不同形态的信息资源，从异构特征空间投影到同形特征空间中，以达到可以直接度量它们之间关系的目的。该过程可作如下形式化描述:

对于给定的信息资源集合 $S=\{S_1,S_2,\cdots,S_k\}$，S_i 在 m 维特征空间 L 中的向量表达为 $S_i=\{L_{S_i}^1,L_{S_i}^2,\cdots,L_{S_i}^m\}$，其在 n 维特征空间 G 中的向量表达为 $S_i=\{G_{S_i}^1,G_{S_i}^2,\cdots,G_{S_i}^n\}$，通过某种策略 F (子空间相关性最大化)或算法，学习到空间投影关系 φ_L 、φ_G 及 t 维特征子空间 O，使得 $\varphi_L(L_{S_i}^1,L_{S_i}^2,\cdots,L_{S_i}^m)\rightarrow(O_{S_i}^1,O_{S_i}^2,\cdots,O_{S_i}^t)$ 、$\varphi_G(G_{S_i}^1,G_{S_i}^2,\cdots,G_{S_i}^n)\rightarrow(O_{S_i}^1,O_{S_i}^2,\cdots,O_{S_i}^t)$，$\varphi_L$ 、φ_G 称为空间投影函数，特征子空间 O 称为最大相关子空间。其几何意义如图 1 所示:

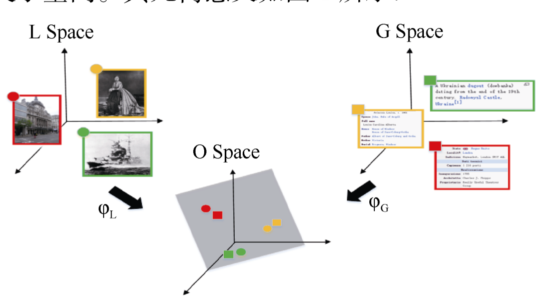


图 1　基于相关性的特征子空间投影示意图

3.3　基于特征子空间的检索排序算法

基于相关性的跨模态信息检索实质上就是在同形特征子空间 O 中，采用某种距离计算方法，度量查询信息资源与被检索信息资源之间的相关性，并按照相关性大小排序。该算法伪代码如下所示:

```
S_to ← φL(St)
for Sj in S do
    S_jo ← φG(Sj)
    Score(St,Sj)=Dis(Sto, Sjo)
end for
Sort S on Score(St, Sj)
```

其中 S_t 为任意查询，其在特征空间 L 中的向量表达为 $S_t=\{L_{S_t}^1,L_{S_t}^2,\cdots,L_{S_t}^m\}$ 。S 为被检索的资源集合，$S_j\in S$，且 S_j 在特征空间 G 中表达。Dis 为距离计算公式，$Score(S_t, S_j)$表示查询 S_t 与记录 S_j 的相关性得分。相似度度量排序的其他策略可直接引用机器学习中的距离计算方法，具体可参见文献[16]。

3.4　偏最小二乘法的应用分析

本文认为基于相关性的跨模态信息检索差异主要在于上述三个核心步骤的不同，即相同步骤下采用不同的策略是导致检索效果差异的主要原因，因此对于任意一个步骤的改进都将有利于提升跨模态信息检索的效果。特征子空间投影是基于相关性的跨模态信息检索研究最核心的步骤，其是现阶段融合不同量纲、不同维度特征数据的唯一途径。目前相关研究[7-8,10-12]多采用典型相关性分析法寻找同一信息在不同形态下的最大相关子投影空间，作为该步骤的执行策略和数学求解方法。然而，典型相关性分析作为一种多元统计分析方法，其利用线性回归表示子投影之间的关系，存在一定的缺陷。

偏最小二乘法作为第二代多元回归分析法，同时兼顾了多元线性回归、主成分分析、典型相关性分析的优点，已被广泛应用于经济学、机械控制技术、社会调查研究、计量化学、神经医学成像等领域。从理论上看，偏最小二乘法不仅能够实现典型相关性分析的功能，还具备去噪音、突出主要潜变量等其他优点，因此本研究认为将偏最小二乘法引入跨模态信息检索框架，将有利于优化基于相关性的跨模态信息检索的结果。偏最小二乘法主要有偏最小二乘回归(PLSR)和偏最小二乘相关(PLSC)两种，前者多用于预测，后者常用于潜变量关联挖掘，具体的数学理论和推导可参

见文献[17]。

基于偏最小二乘法的跨模态信息检索的实质是利用偏最小二乘法(对应于3.2节的策略F和4.1节的特征子空间投影步骤)求解信息从原始特征空间L、G到特征子空间O的映射函数φ_L、φ_G，在保持原始特征之间关联性最大的条件下，突出主成分的作用，抑制数据中的噪音影响。

4 实 验

为探讨偏最小二乘法在跨模态信息检索框架中的应用，本研究设计了相关实验。

4.1 实验数据及相关过程

鉴于语义技术在文本处理和图像分析上的成熟应用，选取文本、图像作为跨模态信息检索的原始信息，以"文本搜图像"和"图像搜文本"两类任务衡量实验的最终结果。实验选用Wikipedia跨模态信息检索数据集[12]，该数据集共包含2 866篇Wikipedia文档和10个主题，每篇文档都由一个"文本–图像"对组成，且属于某一主题；其中2 173篇文档为训练集TRAIN，用于训练空间投影函数φ_L、φ_G和特征子空间O；另外693篇文档为测试集TEST，用于评价跨模态信息检索排序算法的结果，数据分布如表1所示：

表1 数据类型分布表

编号	主题	训练集	测试集	总文档数
#0	艺术与建筑	138	34	172
#1	生物学	272	88	360
#2	地理与位置	244	96	340
#3	历史	248	85	333
#4	文学与戏剧	202	65	267
#5	媒体	178	58	236
#6	音乐	186	51	237
#7	皇室与贵族	144	41	185
#8	体育休闲	214	71	285
#9	战争	347	104	451

依据第3节介绍的基于相关性的跨模态信息检索主体框架，分别对本实验的多模态信息表达、特征子空间投影和相似度度量排序三个核心组成部分作如下说明，其中特征空间L、G、O维度的选择对实验结果影响较小[12]:

(1) 多模态信息表达：对于文档的文本信息，本实验运用gensim①工具包抽取其在LDA主题空间中的特征，构建特征空间L，特征空间L的维度m=10；运用VLFeat②机器视觉库计算其在BOW图像语义空间中的特征，构建特征空间G，特征空间G的维度n=128。

(2) 特征子空间投影：设置三组实验，分别运用scikit-learn③工具包的CCA、PLSR、PLSC三种算法，在训练集数据上学习空间投影函数φ_L、φ_G及文档S_i在特征子空间O中的向量表达$S_i=(O_{S_i}^1, O_{S_i}^2, \cdots, O_{S_i}^t)$，特征子空间维度t=9。

(3) 相似度度量排序：以向量余弦相似度为相关性的度量公式，则文档S_t与文档S_j的相关性得分如下：

$$\mathrm{Score}(S_t,S_j)=\frac{\overline{S_{to}}^T\cdot\overline{S_{jo}}}{\|\overline{S_{to}}\|\times\|\overline{S_{jo}}\|}$$

4.2 实验结果与分析

实验最终在测试集上执行跨模态信息检索，包括"文本搜图像"和"图像搜文本"两个任务，检索相关性判断依据为主题相关，即检索和查询记录主题是否一致。检索结果采用P@K(Precision at K)、MAP(Mean Average Precision)和NDCG (Normalized Discounted Cumulative Gain)三种指标进行评价，从多个角度查看各种方法对检索结果的影响，体现检索结果优化的普适性。

实验对比分析了以CCA(典型相关性分析)、PLSR(偏最小二乘回归)和PLSC(偏最小二乘相关)作为特征子空间学习算法时，"文本搜图像"和"图像搜文本"两个跨模态信息检索任务的P@K(K=5,10,15,20, 30)值，如图2所示。可以看到PLSC方法在两类任务中均获得了最优表现；"文本搜图像"任务中P@K随着K的增大呈递减的趋势，且CCA表示的曲线的斜率较大，而基于偏最小二乘法(PLSR、PLSC)的曲线的斜率较平缓，这表明基于偏最小二乘法所学习的特征子空间投影较CCA方法更稳定；"图像搜文本"任务中三条曲线的斜率都很平缓，与"文本搜图像"任务的表现迥

①http://radimrehurek.com/gensim/.

②http://www.vlfeat.org/.

③http://scikit-learn.org/stable/.

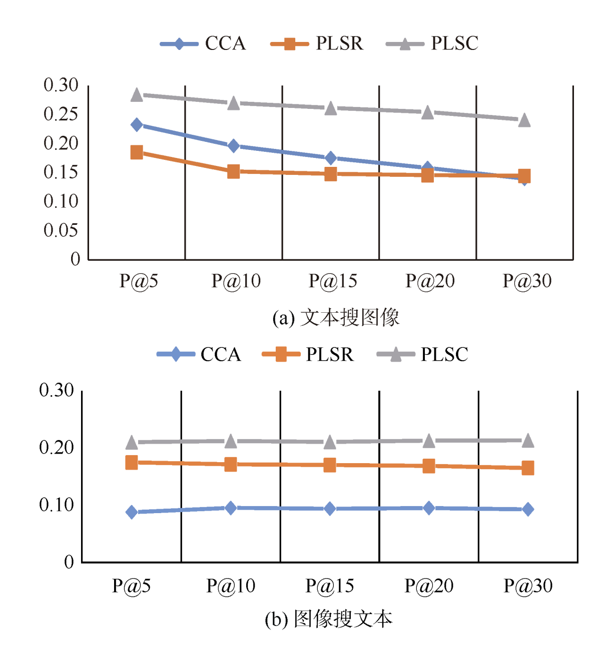


图 2　三种方法的 P@K 对比

异，这说明文本信息在特征子空间中的投影是离散的、均匀的分布，而图像信息在特征子空间中的投影呈明显的按主题聚类特征，如图 3 所示：

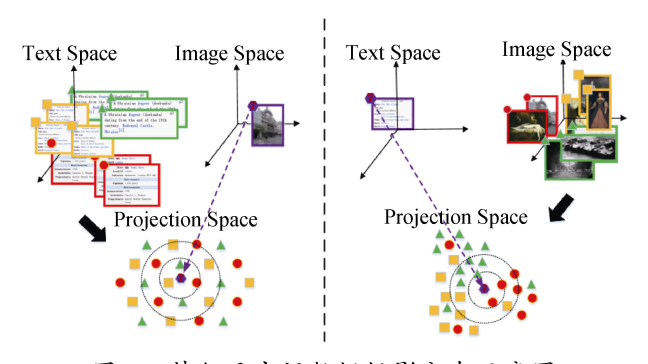


图 3　特征子空间数据投影分布示意图

三组检索实验的 MAP 得分结果如表 2 所示，可以看到 PLSC 方法在两类任务中都得到了最佳效果。与 CCA 方法相比较，“文本搜图像”任务上效果提高了 19.1%，“图像搜文本”任务上效果提高了 36.7%，平均效果提高了 28.2%。通过双尾成对 T 检验可知，两个任务上的效果提升都具有显著性(p1=0.012、p2=0.061)。

表 2　MAP 得分表

方法	文本搜图像	图像搜文本	平均
CCA	0.1645	0.1787	0.1716
PLSR	0.1412	0.1776	0.1594
PLSC	**0.1958**	**0.2443**	**0.2201**

如图 4 所示，比较分析三组实验在两个任务上的 NDCG 的得分可知，从各个主题的 NDCG 得分看，在不同的主题、不同的任务中三种方法的效果各有优劣；从总体 NDCG 得分看，PLSC 在两个任务中均获得最优表现(NDCG 分别为 0.2378 和 0.1982，平均 NDCG 为 0.2179)，该得分相较于 CCA 提高了 70.7%。同样通过双尾成对 T 检验可知，两个任务上的效果提升具有统计显著性意义(p1=0.024、p2=0.036)。PLSR 在“文本搜图像”任务上效果与 CCA 方法基本相同，而在“图像搜文本”任务中效果较 CCA 方法好。

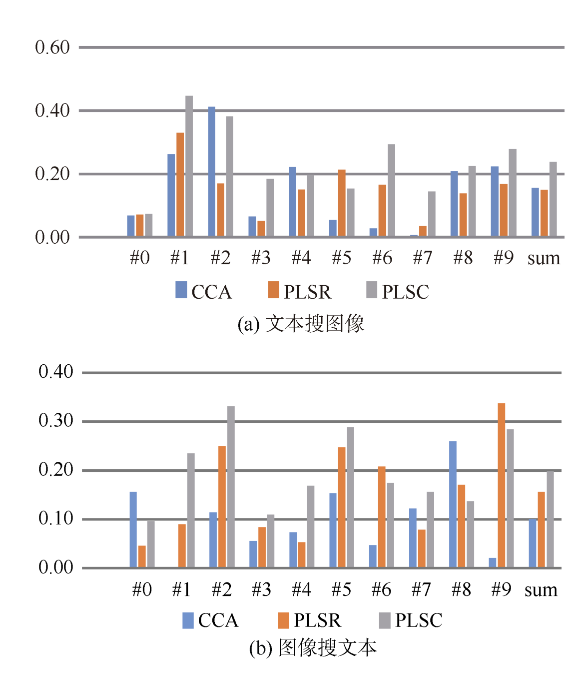


图 4　NDCG 得分簇状柱形图

综合考虑 P@K、MAP 和 NDCG 三个评价指标，偏最小二乘相关法在三种评价指标上效果均优于典型相关性分析法，而偏最小二乘回归法则表现出不稳定的状态，由此认为偏最小二乘相关法更适用于基于相关性的跨模态信息检索理论框架。与典型相关性分析法相比，使用偏最小二乘相关法学习的特征子空间投影与原始空间信息的一致性更强，跨模态信息检索结果更稳定。

5　结　语

基于内容的多媒体信息检索研究日益成熟，“以图搜图”、“哼唱检索”等应用解决了同形态空间内信息资

源的检索问题，但是难以突破异构数据类型的限制。跨模态信息检索研究为该问题提供了一种新的解决思路，然而当前基于相关性的跨模态信息检索研究多运用典型相关性分析法构建特征子空间，存在一定的缺陷。本文将偏最小二乘法引入基于相关性的跨模态信息检索框架，并设计相应的检索实验，实验结果表明偏最小二乘相关算法较好地优化了检索结果。

本文选取文本和图像数据，探讨了偏最小二乘法对这两种不同模态信息资源之间跨媒体相关性的优化，该方法同样适用于其他模态的信息资源(如音频、图像、视频)，以及跨语言信息检索研究。本文的不足之处在于偏最小二乘法在处理数据时假设数据之间的关系是线性的，数据基向量之间是正交关系，因而无法解决非线性、非正交问题。后续研究将聚焦于非线性特征子空间学习，以弥补偏最小二乘法的线性和正交假设所导致的不足。

参考文献：

[1] Smeulders A W M, Worring M, Santini S, et al. Content-based Image Retrieval at the End of the Early Years [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(12): 1349-1380.

[2] Wang A. An Industrial Strength Audio Search Algorithm [C]. In: Proceedings of International Society for Music Information Retrieval Conference, Baltimore, Maryland, USA. 2003: 7-13.

[3] Snoek C G M, Worring M. Concept-based Video Retrieval [J]. Foundations and Trends in Information Retrieval, 2008, 2(4): 215-322.

[4] 王宇新，郭禾，何昌钦，等. 用于图像场景分类的空间视觉词袋模型[J]. 计算机科学，2011，38(8): 265-268. (Wang Yuxin, Guo He, He Changqin, et al. Bag of Spatial Visual Words Model for Scene Classification [J]. Computer Science, 2011, 38(8): 265-268.)

[5] 张志飞，苗夺谦，高灿. 基于 LDA 主题模型的短文本分类方法[J]. 计算机应用，2013，33(6): 1587-1590. (Zhang Zhifei, Miao Duoqian, Gao Can. Short Text Classification Using Latent Dirichlet Allocation [J]. Journal of Computer Applications, 2013, 33(6): 1587-1590.)

[6] 段瑞雪，王小捷，孙月萍，等. HDP 主题模型的用户意图聚类[J]. 北京邮电大学学报，2011，34(S1): 55-58. (Duan Ruixue, Wang Xiaojie, Sun Yueping, et al. Clustering User Goals Based on Hierarchical Dirichlet Process Topic Model [J]. Journal of Beijing University of Posts and Telecommunications, 2011, 34(S1): 55-58.)

[7] Wu F, Zhang H, Zhuang Y. Learning Semantic Correlations for Cross-Media Retrieval [C]. In: Proceedings of IEEE International Conference on Image Processing, Atlanta, USA. IEEE, 2006: 1465-1468.

[8] 张鸿，吴飞，庄越挺. 基于特征子空间学习的跨媒体检索方法[J]. 模式识别与人工智能，2008，21(6): 739-745. (Zhang Hong, Wu Fei, Zhuang Yueting. Cross-Media Retrieval Method Based on Feature Subspace Learning [J]. Pattern Recognition and Artificial Intelligence, 2008, 21(6): 739-745.)

[9] 胡涛，武港山，任桐炜，等. 基于 Ontology 的跨媒体检索技术[J]. 计算机工程, 2009, 35(8): 266-268. (Hu Tao, Wu Gangshan, Ren Tongwei, et al. Ontology-based Cross-media Retrieval Technique [J]. Computer Engineering, 2009, 35(8): 266-268.)

[10] 明均仁，何超. 基于语义关联挖掘的数字图书馆跨媒体检索方法研究[J]. 图书情报工作, 2013, 57(7): 101-105. (Ming Junren, He Chao. Research on Cross-media Retrieval Method in Digital Library Based on Semantic Association Mining [J]. Library and Information Service, 2013, 57(7): 101-105.)

[11] 张鸿. 基于相关性挖掘的跨媒体检索研究[D]. 杭州：浙江大学, 2007. (Zhang Hong. Correlation Mining Based Cross-media Retrieval [D]. Hangzhou: Zhejiang University, 2007.)

[12] Rasiwasia N, Costa Pereira J, Coviello E, et al. A New Approach to Cross-modal Multimedia Retrieval [C]. In: Proceedings of the International Conference on Multimedia. ACM, 2010: 251-260.

[13] Costa Pereira J, Coviello E, Doyle G, et al. On the Role of Correlation and Abstraction in Cross-modal Multimedia Retrieval [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(3): 521-535.

[14] 刘扬，郑逢斌，姜保庆，等. 基于多模态融合和时空上下文语义的跨媒体检索模型的研究[J]. 计算机应用，2009，29(4): 1182-1187. (Liu Yang, Zheng Fengbin, Jiang Baoqing, et al. Research of Cross-media Information Retrieval Model Based on Multimodal fusion and Temporal-spatial Context Semantic [J]. Journal of Computer Applications, 2009, 29(4): 1182-1187.)

[15] Zhai X, Peng Y, Xiao J. Cross-modality Correlation Propagation for Cross-media Retrieval [C]. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan. IEEE, 2012: 2337-2340.

[16] 张宇，刘雨东，计钊. 向量相似度测度方法[J]. 声学技术，2009，28(4): 532-536. (Zhang Yu, Liu Yudong, Ji Zhao. Vector Similarity Measurement Method [J]. Technical Acoustics, 2009, 28(4): 532-536.)

[17] Abdi H, Williams L J. Partial Least Squares Methods: Partial Least Squares Correlation and Partial Least Square Regression [A]. // Methods in Molecular Biology [M]. Humana Press, 2013: 549-579.

作者贡献声明：

丁恒：提出研究命题，设计实施方案，数据分析处理，论文起草与修订；

陆伟：设计研究方案，论文最终版本修订。

收稿日期: 2015-07-06
收修改稿日期: 2015-09-16

A Study on Correlation-based Cross-Modal Information Retrieval

Ding Heng[1] Lu Wei[1,2]
[1](School of Information Management, Wuhan University, Wuhan 430072,China)
[2](Center for the Studies of Information Resources, Wuhan University, Wuhan 430072,China)

Abstract: [**Objective**] Summarize the fundamental strategies and core issues in Cross-Modal Information Retrieval (CMIR) based on correlation, and do research about the pros and cons of using partial least squares in feature subspace projection in order to improve retrieval effect. [**Methods**] Based on Wikipedia CMIR dataset, LDA and BOW models are used as a characteristic expression of text and image resources, cosine distance as the similarity measure, and the least squares method is used to learn subspace projection function replacing canonical correlation analysis method. [**Results**] Using comparative analysis of the influence of three features subspace projection methods named canonical correlation analysis, partial least squares regression, partial least squares correlation on CMIR results according to three retrieval evaluation indicators that are P@K, MAP and NDCG, and the results show that partial least squares correlation obtains the best results. [**Limitations**] In dealing with data, partial least squares method assumes a linear relationship between the data and an orthogonal relationship between the data base vectors, therefore the non-linear, non-orthogonal problem can not be solved. [**Conclusions**] Feature subspace projection learning by using partial least squares correlation is more consistent with original spatial information, and CMIR results are more stable.

Keywords: Cross-Modal Information Retrieval   Partial least squares   Subspace projection